Technical Assignment 4

Charlie Zhang1009855514

Introduction

The dataset is a collection of data from a study on aging and dementia, focusing on the relationship between various demographic, cognitive, and neuroimaging measures and the presence of dementia. It allows for analyses on the impact of demographic factors on dementia progression, the relationship between cognitive function scores and brain volume measures, and the identification of potential biomarkers for dementia through longitudinal and cross-sectional studies.

Data Cleaning and Wrangling

The dataset contains 294 entries and 16 columns, with a mix of numerical and categorical data. The summary statistics for the numerical columns of the dataset provide valuable insights into the distribution of each feature:

Visit: Most subjects have only one visit recorded, with some having up to two visits.

MR Delay: The delay before getting an MRI varies widely among participants, ranging from 0 to 1707 days, with a mean delay of approximately 350 days.

Age: Subjects range from 60 to 98 years old, with a mean age of approximately 76 years. This indicates a study focused on older adults, which is relevant for dementia research.

EDUC: The years of education vary from 6 to 23 years, with an average of approximately 14.5 years.

SES (Socioeconomic Status): Ranges from 1 to 5, with a mean SES of around 2.49, indicating that most subjects are from the middle socioeconomic status.

MMSE (Mini-Mental State Examination): Scores range from 15 to 30, with a mean score of about 27.26. The MMSE score is a critical measure of cognitive function, where lower scores indicate severe cognitive impairment.

CDR (Clinical Dementia Rating): Scores range from 0 to 2, with a mean of 0.30, indicating varying degrees of dementia severity among the subjects.

eTIV (Estimated Total Intracranial Volume), **nWBV** (Normalized Whole Brain Volume), and **ASF** (Atlas Scaling Factor): These brain imaging metrics show variability across subjects, which could be related to the study's focus on brain changes associated with dementia.

There are some missing values in the **SES** (socioeconomic status, 15 missing values) and **MMSE** (Mini-Mental State Examination, 1 missing value) columns that will need to be addressed before analysis. Given the small proportion of missing data, they are replaced with median value to maintain the dataset’s integrity.

Exploratory Data Analysis (EDA)

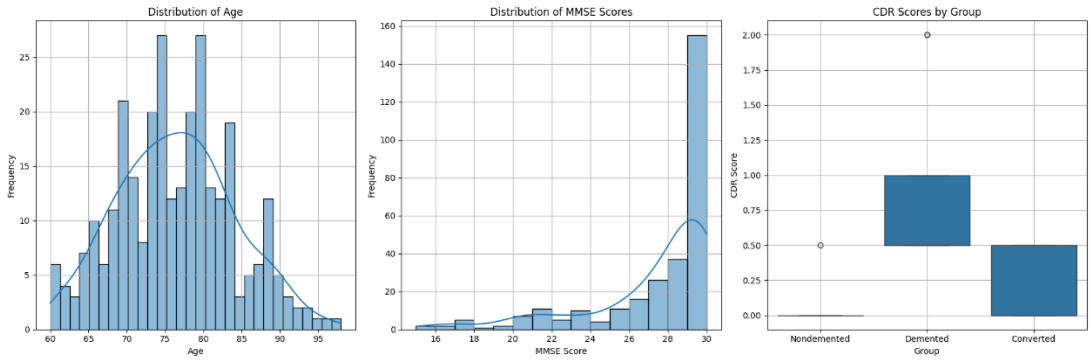

Figure 1: Distribution of Age, MMSE and CDR

Distribution of age, MMSE and CDR (as shown in Figure 1**):**

Age Distribution: The distribution of age among participants is somewhat skewed to the right, indicating a larger number of older participants in the dataset. This is expected in studies related to cognitive impairments or dementia, which are more prevalent in older populations.

MMSE Scores Distribution: The MMSE scores show a left-skewed distribution, with most participants scoring towards the higher end of the scale. This indicates that a large portion of the sample might have normal cognitive function, with fewer participants showing lower scores associated with cognitive impairment.

CDR Scores by Group: The CDR scores, when segmented by group (Demented vs. Nondemented), reveal a clear distinction between the two groups. The 'Demented' group shows higher CDR scores, as expected, indicating more severe cognitive impairment compared to the 'Nondemented' group.

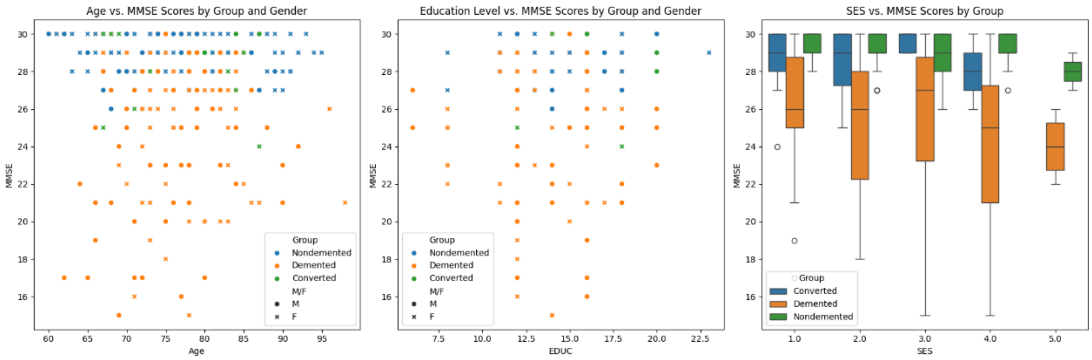

Figure 4: Age, EDUC and SES vs. MMS

Relation in Age, EDUC and SES vs. MMS (as shown in Figure 2**):**

Age vs. MMSE Scores by Group and Gender: Older age brackets show higher participation, typical in dementia research. MMSE scores are dispersed, with the nondemented group predominantly maintaining higher scores, indicating preserved cognitive functions. Gender distribution across age and cognitive groups appears mixed with no overt trend.

Education Level vs. MMSE Scores by Group and Gender: There is a subtle trend of increasing MMSE scores with higher education, especially noticeable in the nondemented group, implying a potential link between education and cognitive health. Gender representation across different education levels is evenly distributed.

SES vs. MMSE Scores by Group: Higher SES is generally associated with better MMSE scores. The nondemented group's scores show less variation by SES, while the demented group's scores decline more with lower SES. The converted group presents intermediate scores, suggesting a shift in cognitive status with varying SES. Outliers are present, indicating score variability, especially among the demented group.

The cognitive scores, assessed by MMSE and CDR, show distinct trajectories over time between individuals with and without dementia. Those with dementia exhibit pronounced cognitive decline, reflected in lower MMSE scores and higher CDR ratings. Demographic and socioeconomic factors play significant roles in this decline, with higher education and SES correlating with better cognitive outcomes, especially evident in the demented group where lower SES exacerbates cognitive decline. Understanding these dynamics informs tailored interventions and support strategies for individuals affected by dementia.

Research Questions:

How do cognitive scores, as measured by the Mini-Mental State Examination (MMSE) and the Clinical Dementia Rating (CDR), change over time in individuals diagnosed with dementia compared to those without dementia, accounting for individual variability?

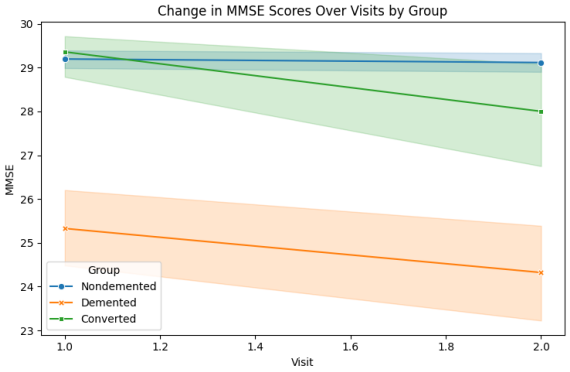

Figure 3: Change in MMSE by group

Figure 4: Change in CDR by group

The MMSE scores graph depicts a decline in cognitive function over two time points, with the 'Demented' group showing a notable decrease in scores, indicative of worsening cognitive abilities. In contrast, the 'Nondemented' group's scores slightly decline, suggesting relatively stable cognitive functions. The presence of a third 'Converted' group implies a transitional cognitive decline, starting from a midpoint and decreasing parallel to the 'Demented' group.

The CDR scores graph illustrates an increasing trend, signifying a progression in cognitive impairment severity over time. The 'Nondemented' group exhibits a gentle upward slope, while the 'Demented' group's scores rise more steeply, highlighting a more rapid cognitive decline. The 'Converted' group's trajectory suggests a shift towards greater impairment, starting near the

'Nondemented' baseline but escalating towards the 'Demented' group's level.

Term	Coefficient	Standard Error	z-Value	P-Value	95% lower	95% upper
Intercept	29.357	0.606	48.418	<0.001	28.169	30.546
Group[T.Demented]	-4.029	0.669	-6.019	<0.001	-5.341	-2.717
Group[T.Nondemented]	-0.163	0.663	-0.246	0.806	-1.461	1.136
Visit[T.2]	-1.336	0.653	-2.046	0.041	-2.616	-0.056
Group[T.Demented]:Visit[T.2]	0.422	0.713	0.591	0.555	-0.977	1.820
Group[T.Nondemented]:Visit[T.2]	1.264	0.707	1.788	0.074	-0.122	2.649

Term	Coefficient	Standard Error	z-Value	P-Value	95% CI Lower	95% CI Upper
Intercept	0.036	0.037	0.964	0.335	-0.037	0.108
Group[T.Demented]	0.566	0.041	13.842	<0.001	0.486	0.646
Group[T.Nondemented]	-0.036	0.040	-0.882	0.378	-0.115	0.044
Visit[T.2]	0.292	0.063	4.601	<0.001	0.168	0.416
Group[T.Demented]:Visit[T.2]	-0.171	0.069	-2.464	0.014	-0.307	-0.035
Group[T.Nondemented]:Visit[T.2]	-0.285	0.069	-4.147	<0.001	-0.420	-0.150

For the MMSE model:

The intercept is highly significant, indicating a high baseline MMSE score. 'Demented' group shows a significantly lower MMSE score compared to the baseline, indicating poorer cognitive function. The coefficient for the second visit is significant, showing a decrease in MMSE scores over time. However, the interactions between group and visit are not significant, suggesting the rate of change over time is not differentially impacting the groups in this model

For the CDR model:

The intercept is not significant, suggesting no baseline difference in CDR scores when all other variables are at zero. 'Demented' group has a significantly higher CDR score, indicating more severe cognitive impairment. The increase in CDR scores from the first to the second visit is significant, indicating progression of impairment over time. The negative coefficients for the interactions indicate that the rate of increase in CDR scores is less for the subsequent visit for both groups, with a more pronounced effect for the 'Nondemented' group.

The results from these models are telling a consistent story of cognitive decline over time, especially among the 'Demented' group. The interaction terms suggest that the rate of decline in MMSE scores and the rate of increase in CDR scores over time might not be as straightforward and could be influenced by other factors not accounted for in the models.

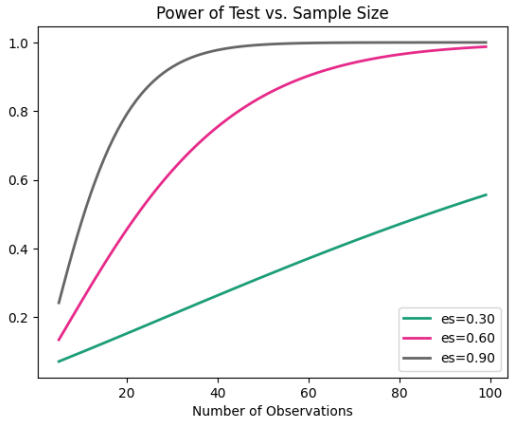

Figure 5: Power analysis

The power analysis graph illustrates how statistical power increases with sample size for three different effect sizes. To achieve a power of 0.91 with an alpha of 0.05 for an effect size of 0.7, the required sample size per group is approximately 45.451. This means that for a t-test to detect a medium-to-large effect with a high probability of correctly rejecting the null hypothesis (high power), a total sample of at least 91 (approximately 46 participants per group when considering two groups) would be needed.

Conclusion

In conclusion, the comprehensive analysis of the dataset sheds light on the intricate relationship between demographic factors, cognitive scores, and dementia progression. The exploration of cognitive trajectories over time, particularly through MMSE and CDR assessments, elucidates the distinct patterns observed in individuals with and without dementia. Moreover, the influence of demographic and socioeconomic factors on cognitive decline underscores the importance of tailored interventions and support strategies for affected individuals. The statistical models further reinforce these findings, emphasizing the significant impact of dementia on cognitive function and the nuanced nature of decline over time. Additionally, the power analysis highlights the importance of adequate sample sizes for robust statistical inference, particularly in studies examining cognitive outcomes. Collectively, these insights provide valuable groundwork for further research aimed at understanding and addressing the complexities of dementia and cognitive decline in aging populations.